Indexed Fields for Bulk Curation of Collection Metadata in CMR (Recommendation)

2015-07-14

I. Summary:

In order to perform bulk curation (updates to more than one record at a time) of collection metadata in CMR, the UMM fields need to be indexed. By indexing the UMM fields, they will be searchable and therefore can be queried and content can be updated. The decision to initially index only the fields below was based on search and not curation and that there would be a curation cache. Metadata curators need the functionality to update any field in order to make timely updates and ensure metadata quality. If the bulk curation functionality is not available, metadata curation maintenance will significantly increase (e.g.: one-by-one update of metadata) and there will be more room for error.

Key Points:
1. Errors in metadata exist in more than just the currently searched fields listed in Section II below.
2. Metadata in those fields are important to users because they contain information that help the user search, understand, and use data.
3. We have identified the critical fields that are not indexed, but are important to users in section III.
4. We request that those fields be indexed and made searchable for curation purposes.
5. If the fields in section III below are not indexed, metadata quality will potentially degrade.

II. Fields Currently Indexed in the CMR:

- Concept ID
- Provider ID
- Entry title
- Data type
- Short name
- Long name
- Summary
- Version ID
- Version description
- Processing level ID
- Science keywords
- Archive centers
- Additional attribute names, data types, values, and descriptions
- Spatial keywords
- Temporal keywords
- Associated DIFs
- Project short and long names
- Platform short and long names
- Instrument short names, long names, and techniques
- Sensor short names, long names, and techniques
- Characteristic names and descriptions
- TwoD coordinate system names
- Collection Data Type

III. Additional Fields That Need to Be Indexed To Perform Bulk Curation:

Most Commonly Updated Fields:
1. Personnel
2. Related URL/URL
3. Privacy Status
4. IDN Node
5. Temporal Coverage
6. Dataset Citation
7. Related URL Content Types
8. Dataset Progress
9. Horizontal Resolution Range
10. Temporal Resolution Range
11. Vertical Resolution Range
12. Distribution
13. Reference Publication
14. Extended Metadata
15. Multimedia Sample

Less Commonly Updated Fields:
1. Quality
2. Access Constraints
3. Use Constraints
4. Ancillary Keyword
5. Related URL/Description
6. Spatial Coverage
7. Parent Metadata
8. Dataset Language
9. Originating Center
10. Multimedia Sample
11. Metadata Association
12. Additional Attributes